Combined associations of a polygenic risk score and classical risk factors with breast cancer risk

Title - Combined associations of a polygenic risk score and classical risk factors with breast cancer risk

Authors and Affiliations:

Pooja Middha Kapoor, M.P.H.[1, 2], Nasim Mavaddat, Ph.D.[3], Parichoy Pal Choudhury, Ph.D.[4], Amber N. Wilcox, M.P.H. [4, 5], Sara Lindström, Ph.D. [6, 7], Sabine Behrens, Ph.D. [1], Kyriaki Michailidou, Ph.D.[3, 8], Joe Dennis, M.Sc.[3], Manjeet K. Bolla, M.Sc.[3], Qin Wang, M.Sc.[3], Audrey Jung, Ph.D.[1], Zomoroda Abu-Ful, M.D.[9], Thomas Ahearn, Ph.D.[4], Irene L. Andrulis, Ph.D.[10, 11], Hoda Anton-Culver, Ph.D.[12], Volker Arndt, M.D.[13], Kristan J. Aronson, Ph.D.[14], Paul L. Auer, Ph.D.[15, 16], Laura E. Beane Freeman, Ph.D.[4], Heiko Becher, Ph.D.[17, 18], Matthias W. Beckmann, M.D.[19], Alicia Beeghly-Fadiel, Ph.D.[20], Javier Benitez, Ph.D.[21, 22], Leslie Bernstein, Ph.D.[23], Stig E. Bojesen,Dr.Med.Sci.[24-26], Hiltrud Brauch, Ph.D.[27-29], Hermann Brenner, M.D.[13, 29, 30], Thomas Brüning, M.D.[31], Qiuyin Cai, M.D.[20], Daniele Campa, Ph.D.[1, 32], Federico Canzian, Ph.D.[33], Angel Carracedo, M.D.[34, 35], Brian D. Carter, M.P.H.[36], Jose E. Castelao, M.D.[37], Stephen J. Chanock, M.D.[4], Nilanjan Chatterjee, Ph.D.[4, 38, 39], Georgia Chenevix-Trench, Ph.D.[40], Christine L. Clarke, Ph.D.[41], Fergus J. Couch, Ph.D.[42], Angela Cox, Ph.D.[43], Simon S. Cross, M.D.[44], Kamila Czene, Ph.D.[45], James Y. Dai, Ph.D.[15], H. Shelton Earp, M.D.[46], Arif B. Ekici, Ph.D.[47], A. Heather Eliassen,Sc.D.[48, 49], Mikael Eriksson, M.Sc.[45], D. Gareth Evans, M.D.[50, 51], Peter A. Fasching, M.D.[19, 52], Jonine Figueroa, Ph.D.[4, 53, 54], Lin Fritschi, M.B.B.S.[55], Marike Gabrielson, Ph.D.[45], Manuela Gago-Dominguez, M.D.[34, 56], Chi Gao, Sc.D.[49, 57], Susan M. Gapstur, Ph.D.[36], Mia M. Gaudet, Ph.D.[36], Graham G. Giles, Ph.D.[58-60], Anna González-Neira, Ph.D.[22], Pascal Guénel, Ph.D.[61], Lothar Haeberle, Ph.D.[62], Christopher A. Haiman, Sc.D.[63], Niclas Håkansson, Ph.D.[64], Per Hall, Ph.D.[45, 65], Ute Hamann, Ph.D.[66], Sigrid Hatse, Ph.D.[67], Jane Heyworth, Ph.D.[68], Bernd Holleczek, Ph.D.[69], Robert N. Hoover, M.D.[4], John L. Hopper, Ph.D.[59], Anthony Howell, M.B.B.S.[70], David J. Hunter, Sc.D.[49, 57, 71], ABCTB

Investigators[72], kConFab/AOCS Investigators[73, 74], Esther M. John, Ph.D.[75], Michael E. Jones, Ph.D.[76], Rudolf Kaaks, Ph.D.[1], Renske Keeman, M.Sc.[77], Cari M. Kitahara, Ph.D.[78], Yon-Dschun Ko, M.D.[79], Stella Koutros, Ph.D.[4], Allison W. Kurian, M.D.[75, 80], Diether Lambrechts, Ph.D.[81, 82], Loic Le Marchand, M.D.[83], Eunjung Lee, Ph.D.[63], Flavio Lejbkowicz, Ph.D.[9], Martha Linet, M.D.[78], Jolanta Lissowska, Ph.D.[84], Ana Llaneza, Ph.D.[85], Robert J. MacInnis, Ph.D.[58, 59], Maria Elena Martinez, Ph.D.[56, 86], Tabea Maurer, Dipl. Psyc.[87], Catriona McLean, M.D.[88], Susan L. Neuhausen, Ph.D.[23], William G. Newman, Ph.D.[89, 90], Aaron Norman, Ph.D.[91], Katie M. O'Brien, M.S.P.H.[92], Andrew F. Olshan, Ph.D.[93], Janet E. Olson, Ph.D.[91], Håkan Olsson, Ph.D.[94], Nick Orr, Ph.D.[95], Charles M. Perou, Ph.D.[96], Guillermo Pita, Ph.D.[97], Eric C. Polley, Ph.D.[91], Ross L. Prentice, Ph.D.[15], Gad Rennert, M.P.H.[9], Hedy S. Rennert, M.D.[9], Kathryn J. Ruddy, M.D.[98], Dale P. Sandler, Ph.D.[92], Christobel Saunders, M.B.B.S.[99], Minouk J. Schoemaker, Ph.D.[76], Ben Schöttker, Ph.D.[13, 100], Fredrick Schumacher, Ph.D.[101], Christopher Scott, M.S.[91], Rodney J. Scott, Ph.D.[102-104], Xiao-Ou Shu, M.D.[20], Ann Smeets, Ph.D.[105], Melissa C. Southey, Ph.D.[60, 106, 107], John J. Spinelli, Ph.D.[108, 109], Jennifer Stone, Ph.D.[59, 110], Anthony J. Swerdlow, D.Sc.[76, 111], Rulla M. Tamimi, Sc.D.[48, 49, 57], Jack A. Taylor, Ph.D.[92, 112], Melissa A. Troester, Ph.D.[93], Celine M. Vachon, Ph.D.[113], Elke M. van Veen, M.Sc.[89, 90], Xiaoliang Wang, Ph.D.[6, 7], Clarice R. Weinberg, Ph.D.[114], Caroline Weltens, Ph.D.[115], Walter Willett, Dr.P.H.[49, 116, 117], Stacey J. Winham, Ph.D.[118], Alicja Wolk, Dr.Med.Sci.[64, 119], Xiaohong R. Yang, Ph.D.[4], Wei Zheng, Ph.D.[20], Argyrios Ziogas, Ph.D.[12], Alison M. Dunning, Ph.D.[120], Paul D.P. Pharoah, Ph.D.[3, 120], Marjanka K. Schmidt, Ph.D.[77, 121], Peter Kraft, Ph.D.[49, 57], Douglas F. Easton, Ph.D.[3, 120], Roger L. Milne, Ph.D.[58-60], Montserrat García-Closas, Dr.P.H.[4] †, Jenny Chang-Claude, Ph.D.[1, 87] †*

1. Division of Cancer Epidemiology. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
2. Faculty of Medicine. *University of Heidelberg*. Heidelberg: Germany; 69120.
3. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care. *University of Cambridge*. Cambridge: UK; CB1 8RN.

4. Division of Cancer Epidemiology and Genetics. *National Cancer Institute, National Institutes of Health, Department of Health and Human Services*. Bethesda, MD: USA; 20850.
5. Department of Epidemiology, Gillings School of Global Public Health. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
6. Department of Epidemiology. *University of Washington School of Public Health*. Seattle, WA: USA; 98195.
7. Public Health Sciences Division. *Fred Hutchinson Cancer Research Center*. Seattle, WA: USA; 98109.
8. Department of Electron Microscopy/Molecular Pathology and The Cyprus School of Molecular Medicine. *The Cyprus Institute of Neurology & Genetics*. Nicosia: Cyprus; 1683.
9. Clalit National Cancer Control Center. *Carmel Medical Center and Technion Faculty of Medicine*. Haifa: Israel; 35254.
10. Fred A. Litwin Center for Cancer Genetics. *Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital*. Toronto, ON: Canada; M5G 1X5.
11. Department of Molecular Genetics. *University of Toronto*. Toronto, ON: Canada; M5S 1A8.
12. Department of Epidemiology, Genetic Epidemiology Research Institute. *University of California Irvine*. Irvine, CA: USA; 92617.
13. Division of Clinical Epidemiology and Aging Research, C070. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
14. Department of Public Health Sciences, and Cancer Research Institute. *Queen's University*. Kingston, ON: Canada; K7L 3N6.
15. Cancer Prevention Program. *Fred Hutchinson Cancer Research Center*. Seattle, WA: USA; 98109.
16. Zilber School of Public Health. *University of Wisconsin-Milwaukee*. Milwaukee, WI: USA; 53205.
17. Institute of Medical Biometry and Epidemiology. *University Medical Center Hamburg-Eppendorf*. Hamburg: Germany; 20246.
18. Institute of Biometry and Clinical Epidemiology. *Charité - Universitätsmedizin Berl* Berlin: Germany; 10117.
19. Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN. *University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg*. Erlangen: Germany; 91054.
20. Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center. *Vanderbilt University School of Medicine*. Nashville, TN: USA; 37232.
21. *Centro de Investigación en Red de Enfermedades Raras (CIBERER)*. Madrid: Spain; 28029.
22. Human Cancer Genetics Programme. *Spanish National Cancer Research Centre (CNIO)*. Madrid: Spain; 28029.
23. Department of Population Sciences. *Beckman Research Institute of City of Hope*. Duarte, CA: USA; 91010.
24. Copenhagen General Population Study, Herlev and Gentofte Hospital. *Copenhagen University Hospital*. Herlev: Denmark; 2730.
25. Department of Clinical Biochemistry, Herlev and Gentofte Hospital. *Copenhagen University Hospital*. Herlev: Denmark; 2730.
26. Faculty of Health and Medical Sciences. *University of Copenhagen*. Copenhagen: Denmark; 2200.
27. *Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology*. Stuttgart: Germany; 70376.
28. iFIT-Cluster of Excellence. *University of Tuebingen*. Tuebingen: Germany; 72074.
29. German Cancer Consortium (DKTK). *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
30. Division of Preventive Oncology. *German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT)*. Heidelberg: Germany; 69120.

31. Institute for Prevention and Occupational Medicine of the German Social Accident Insurance. *Institute of the Ruhr University Bochum (IPA)*. Bochum: Germany; 44789.
32. Department of Biology. *University of Pisa*. Pisa: Italy; 56126.
33. Genomic Epidemiology Group. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
34. Genomic Medicine Group, Galician Foundation of Genomic Medicine. *Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS*. Santiago de Compostela: Spain; 15706.
35. Centro de Investigación en Red de Enfermedades Raras (CIBERER) y Centro Nacional de Genotipado (CEGEN-PRB2). *Universidad de Santiago de Compostela*. Santiago De Compostela: Spain; 15782.
36. Behavioral and Epidemiology Research Group. *American Cancer Society*. Atlanta, GA: USA; 30303.
37. Oncology and Genetics Unit. *Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS*. Vigo: Spain; 36312.
38. Department of Biostatistics, Bloomberg School of Public Health. *John Hopkins University*. Baltimore, MD: USA; 21205.
39. Department of Oncology, School of Medicine. *John Hopkins University*. Baltimore, MD: USA; 21205.
40. Department of Genetics and Computational Biology. *QIMR Berghofer Medical Research Institute*. Brisbane, Queensland: Australia; 4006.
41. Westmead Institute for Medical Research. *University of Sydney*. Sydney, New South Wales: Australia; 2145.
42. Department of Laboratory Medicine and Pathology. *Mayo Clinic*. Rochester, MN: USA; 55905.
43. Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism. *University of Sheffield*. Sheffield: UK; S10 2TN.
44. Academic Unit of Pathology, Department of Neuroscience. *University of Sheffield*. Sheffield: UK; S10 2TN.
45. Department of Medical Epidemiology and Biostatistics. *Karolinska Institutet*. Stockholm: Sweden; 171 65.
46. Lineberger Comprehensive Cancer Center. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
47. Institute of Human Genetics, University Hospital Erlangen. *Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN*. Erlangen: Germany; 91054.
48. Channing Division of Network Medicine, Department of Medicine. *Brigham and Women's Hospital and Harvard Medical School*. Boston, MA: USA; 02115.
49. Department of Epidemiology. *Harvard T.H. Chan School of Public Health*. Boston, MA: USA; 02115.
50. Genomic Medicine, Division of Evolution and Genomic Sciences. *The University of Manchester, Manchester Academic Health Science Centre, Manchester Universities Foundation Trust, St. Mary's Hospital*. Manchester: UK; M13 9WL.
51. Genomic Medicine, North West Genomics hub. *Manchester Academic Health Science Centre, Manchester Universities Foundation Trust, St. Mary's Hospital*. Manchester: UK; M13 9WL.
52. David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology. *University of California at Los Angeles*. Los Angeles, CA: USA; 90095.
53. Usher Institute of Population Health Sciences and Informatics. *The University of Edinburgh Medical School*. Edinburgh: UK; EH16 4TJ.
54. *Cancer Research UK Edinburgh Centre*. Edinburgh: UK; EH4 2XR.
55. School of Public Health. *Curtin University*. Perth, Western Australia: Australia; 6102.
56. Moores Cancer Center. *University of California San Diego*. La Jolla, CA: USA; 92037.

57. Program in Genetic Epidemiology and Statistical Genetics. *Harvard T.H. Chan School of Public Health*. Boston, MA: USA; 02115.
58. Cancer Epidemiology Division. *Cancer Council Victoria*. Melbourne, Victoria: Australia; 3004.
59. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health. *The University of Melbourne*. Melbourne, Victoria: Australia; 3010.
60. Precision Medicine, School of Clinical Sciences at Monash Health. *Monash University*. Clayton, Victoria: Australia; 3168.
61. Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP). *INSERM, University Paris-Sud, University Paris-Saclay*. Villejuif: France; 94805.
62. Department of Gynaecology and Obstetrics, University Hospital Erlangen. *Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN*. Erlangen: Germany; 91054.
63. Department of Preventive Medicine, Keck School of Medicine. *University of Southern California*. Los Angeles, CA: USA; 90033.
64. Institute of Environmental Medicine. *Karolinska Institutet*. Stockholm: Sweden; 171 77.
65. Department of Oncology. *Södersjukhuset*. Stockholm: Sweden; 118 83.
66. Molecular Genetics of Breast Cancer. *German Cancer Research Center (DKFZ)*. Heidelberg: Germany; 69120.
67. Laboratory of Experimental Oncology (LEO), Department of Oncology. *Leuven Cancer Institute, KU Leuven*. Leuven: Belgium; 3000.
68. School of Population and Global Health. *The University of Western Australia*. Perth, Western Australia: Australia; 6009.
69. *Saarland Cancer Registry*. Saarbrücken: Germany; 66119.
70. Division of Cancer Sciences. *University of Manchester*. Manchester: UK; M13 9PL.
71. Nuffield Department of Population Health. *University of Oxford*. Oxford: UK; OX3 7LF.
72. Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research. *University of Sydney*. Sydney, New South Wales: Australia; 2145.
73. Research Department. *Peter MacCallum Cancer Center*. Melbourne, Victoria: Australia; 3000.
74. Sir Peter MacCallum Department of Oncology. *The University of Melbourne*. Melbourne, Victoria: Australia; 3000.
75. Department of Medicine, Division of Oncology and Stanford Cancer Institute. *Stanford University School of Medicine*. Stanford, CA: USA; 94304.
76. Division of Genetics and Epidemiology. *The Institute of Cancer Research*. London: UK; SM2 5NG.
77. Division of Molecular Pathology. *The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital*. Amsterdam: The Netherlands; 1066 CX.
78. Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics. *National Cancer Institute*. Bethesda, MD: USA; 20892.
79. Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH. *Johanniter Krankenhaus*. Bonn: Germany; 53177.
80. Department of Health Research and Policy - Epidemiology. *Stanford University School of Medicine*. Stanford, CA: USA; 94305.
81. VIB Center for Cancer Biology. *VIB*. Leuven: Belgium; 3001.
82. Laboratory for Translational Genetics, Department of Human Genetics. *University of Leuven*. Leuven: Belgium; 3000.
83. Epidemiology Program. *University of Hawaii Cancer Center*. Honolulu, HI: USA; 96813.
84. Department of Cancer Epidemiology and Prevention. *M. Sklodowska-Curie Cancer Center, Oncology Institute*. Warsaw: Poland; 02-034.
85. General and Gastroenterology Surgery Service. *Hospital Universitario Central de Asturias*. Oviedo: Spain; 33011.

86. Department of Family Medicine and Public Health. *University of California San Diego*. La Jolla, CA: USA; 92093.
87. Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH). *University Medical Center Hamburg-Eppendorf*. Hamburg: Germany; 20246.
88. Anatomical Pathology. *The Alfred Hospital*. Melbourne, Victoria: Australia; 3004.
89. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health. *University of Manchester, Manchester Academic Health Science Centre*. Manchester: UK; M13 9WL.
90. Manchester Centre for Genomic Medicine. *St Mary's Hospital, Manchester NIHR Biomedical Research Centre, Manchester University Hospitals NHS, Foundation Trust, Manchester Academic Health Science Centre*. Manchester: UK; M13 9WL.
91. Department of Health Sciences Research. *Mayo Clinic*. Rochester, MN: USA; 55905.
92. Epidemiology Branch. *National Institute of Environmental Health Sciences, NIH*. Research Triangle Park, NC: USA; 27709.
93. Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
94. Department of Cancer Epidemiology, Clinical Sciences. *Lund University*. Lund: Sweden; 222 42.
95. Centre for Cancer Research and Cell Biology. *Queen's University Belfast*. Belfast, Ireland: UK; BT7 1NN.
96. Department of Genetics, Lineberger Comprehensive Cancer Center. *University of North Carolina at Chapel Hill*. Chapel Hill, NC: USA.
97. Human Genotyping-CEGEN Unit, Human Cancer Genetic Program. *Spanish National Cancer Research Centre*. Madrid: Spain; 28029.
98. Department of Oncology. *Mayo Clinic*. Rochester, MN: USA; 55905.
99. School of Medicine. *University of Western Australia*. Perth, Western Australia: Australia.
100. Network Aging Research. *University of Heidelberg*. Heidelberg: Germany; 69115.
101. Department of Population and Quantitative Health Sciences. *Case Western Reserve University*. Cleveland, OH: USA; 44106.
102. Division of Molecular Medicine, Pathology North. *John Hunter Hospital*. Newcastle, New South Wales: Australia; 2305.
103. Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health. *University of Newcastle*. Callaghan, New South Wales: Australia; 2308.
104. Hunter Medical Research Institute. *John Hunter Hospital*. Newcastle, New South Wales: Australia; 2305.
105. Department of Surgical Oncology. *University Hospitals Leuven*. Leuven: Belgium; 3000.
106. Department of Clinical Pathology. *The University of Melbourne*. Melbourne, Victoria: Australia; 3010.
107. Cancer Epidemiology Centre. *Cancer Council Victoria*. Melbourne, Victoria: Australia; 3004.
108. Population Oncology. *BC Cancer*. Vancouver, BC: Canada; V5Z 1G1.
109. School of Population and Public Health. *University of British Columbia*. Vancouver, BC: Canada; V6T 1Z4.
110. The Curtin UWA Centre for Genetic Origins of Health and Disease. *Curtin University and University of Western Australia*. Perth, Western Australia: Australia; 6000.
111. Division of Breast Cancer Research. *The Institute of Cancer Research*. London: UK; SW7 3RP.
112. Epigenetic and Stem Cell Biology Laboratory. *National Institute of Environmental Health Sciences, NIH*. Research Triangle Park, NC: USA; 27709.
113. Department of Health Science Research, Division of Epidemiology. *Mayo Clinic*. Rochester, MN: USA; 55905.
114. Biostatistics and Computational Biology Branch. *National Institute of Environmental Health Sciences, NIH*. Research Triangle Park, NC: USA; 27709.

115. Leuven Multidisciplinary Breast Center, Department of Oncology. *Leuven Cancer Institute, University Hospitals Leuven*. Leuven: Belgium; 3000.
116. Department of Nutrition. *Harvard T.H. Chan School of Public Health*. Boston, MA: USA; 02115.
117. Channing Division of Network Medicine. *Brigham and Women's Hospital and Harvard Medical School*. Boston, MA: USA; 02115.
118. Department of Health Science Research, Division of Biomedical Statistics and Informatics. *Mayo Clinic*. Rochester, MN: USA; 55905.
119. Department of Surgical Sciences. *Uppsala University*. Uppsala: Sweden; 751 05.
120. Centre for Cancer Genetic Epidemiology, Department of Oncology. *University of Cambridge*. Cambridge: UK; CB1 8RN.
121. Division of Psychosocial Research and Epidemiology. *The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital*. Amsterdam: The Netherlands; 1066 CX.

Keywords: Polygenic risk score, Breast cancer, Risk factors, Genetic susceptibility, Gene-environment interactions, Risk prediction, Epidemiology

Abstract – 192 words

Main text – 1184 words

*Corresponding author.

†These authors contributed equally to this work.

Corresponding author:

Jenny Chang-Claude

Unit of Genetic Epidemiology, Division of Cancer Epidemiology

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 581, 69120 Heidelberg, Germany

Tel: +49-6221 42 2373, Fax: +49-6221 42 2203

E-mail: j.chang-claude@dkfz.de

Abstract

Combining polygenic risk score (PRS) with classical risk factors could improve accuracy of breast cancer prediction models. We evaluated the joint associations between a new 313-variant PRS (PRS_{313}) and questionnaire-based breast cancer risk factors in women of European ancestry, using 72,284 cases and 80,354 controls from the Breast Cancer Association Consortium. Interactions were evaluated using standard logistic regression, and a newly developed case-only method, for breast cancer risk overall and by ER status. After accounting for multiple testing, we did not find evidence that per-standard deviation PRS_{313} odds ratio differed across strata defined by individual risk factors. Global and tail-based goodness of fit tests also did not reject the assumption of a multiplicative model between PRS_{313} and each risk factor. Variation in projected absolute lifetime risk of breast cancer from classical risk factors was greater for women with higher genetic risk (PRS_{313} and family history). The average absolute risk difference for women in the lowest and highest deciles of genetic risk due to all classical and the subset of modifiable risk factors is 17.5% and 16.5%, respectively. These findings have implications for development of comprehensive models for risk-stratified prevention of breast cancer.

Main text

Precision prevention and early-detection of cancer is a key focus of cancer research and utilizes tools such as risk prediction models for risk stratification[1, 2]. Many breast cancer risk prediction models are focused either on classical risk factors or on inherited mutations conferring moderate-to-high risk of cancer, and do not include risk conferred by common susceptibility variants[3]. Modeling the joint associations of genetic and classical risk factors could result in substantial improvement in risk stratification and therefore improved prevention and screening modalities for breast cancer[4-7] .

Combined associations of SNPs can be summarized by a polygenic risk score (PRS); women in the top 1% of the newly derived 313-SNP PRS (PRS_{313}) have a four-fold increased risk of breast cancer relative to women at population-average risk[8]. Previous studies, which evaluated combined associations between classical risk factors and breast cancer PRS based on 77 SNPs[9] and 24 SNPs[10], found weak or no evidence of departure from the multiplicative risk model for overall breast cancer. In the current study, we extend these analyses to assess the combined associations of the PRS_{313} and classical risk factors using data from the Breast Cancer Association Consortium (BCAC). This new PRS has been validated in prospective studies and shown to be more predictive than the previously reported 77-SNP PRS[11] for risk of overall breast cancer as well as for estrogen receptor (ER) subtype-specific breast cancer[8]. Additionally, this study found evidence of interaction between PRS_{313} and family history for ER-positive disease, indicating the need to consider the joint effects of PRS_{313} and family history[8].

Detailed information on study population, genetic data and risk factor data is provided in the **Supplementary Materials**. Briefly, we performed analyses in women of European ancestry,

using data from 16 prospective cohorts, 14 population-based case-control studies and 16 non-population based studies participating in BCAC **(Supplementary Table 1)**. Samples were genotyped using two arrays, iCOGS[12] and OncoArray[13-15]. Data on risk factors were derived with respect to a reference age (date at diagnosis for cases and date at interview for controls).

The development of the PRS is briefly explained in Supplementary Materials[8]. The individual PRS was derived using a weighted sum of genotypes of the form:

$$PRS = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k .... + \beta_n x_n$$

where β_k is per-allele log odds ratio for breast cancer associated with minor allele of SNP k, x_k is the dosage of the allele for SNP k and n is the total number of SNPs included in the PRS. Subtype-specific PRSs were created by incorporating ER subtype-specific weights[8]. We standardized the PRS to have unit standard deviation of controls.

Departure from multiplicative model between standardized PRS_{313} and each of the risk factors were assessed using two methods, (i) unconditional logistic regression models and likelihood ratio tests and (ii) a newly developed case-only method, which assumes independence between PRS and risk factors in the underlying population and shows greater efficiency compared to logistic regression[16]. Individual models were fitted for each PRS-risk factor combination for overall and ER-specific breast cancer. ER-specific PRS were used for interaction analyses of the corresponding ER-specific breast cancer risk. Models were adjusted for reference age, study, and corresponding ten ancestry-informative principal components for each array. Array-specific results were meta-analyzed using a fixed-effect inverse-variance weighted method. To evaluate global goodness-of-fit of the multiplicative model between PRS_{313}

and for each risk factor, we performed the Hosmer-Lemeshow test using population-based studies. Moreover, we assessed goodness-of-fit at the extremes of the distribution (tails) using the tail-based test[17]. Using the iCARE-BPC3 model[4], we projected absolute lifetime risk of breast cancer for 50-year old White non-Hispanic US women before 80 years. We assessed the distribution of risk due to classical risk factors (i.e. menstrual/reproductive, and lifestyle) within categories of risk defined by genetic factors (i.e. breast cancer family history and PRS_{313}).

Associations between PRS_{313} and overall and ER-specific breast cancer risk are higher than expected due to over-fitting as there is substantial overlap between the SNP discovery population and our dataset (**Supplementary Figure 1)**. The number of cases and controls varied for each risk factor, ranging from 61,617 cases and 74,698 controls for ever parous to 14,576 cases and 19,640 controls for pack-years smoked for overall breast cancer risk **(Supplementary Table 2)**. Based on the population-based case-control and prospective cohort studies, the associations of the risk factors with overall and ER subtype-specific breast cancer were of the expected magnitude and direction **(Supplementary Table 3)**.

After accounting for multiple testing using Bonferroni correction ($p_{int} < 0.05/16 = 0.003$), none of the interactions between PRS_{313} and the classical risk factors were statistically significant except for family history **(Table 1)**. The interaction between PRS_{313} and family history for ER-positive breast disease is in agreement with what has been previously published [8]. Further, we also observed an interaction for overall and ER-negative breast cancer risk.

There was no clear dose-response pattern in the estimated ORs associated with classical risk factors stratified by PRS percentiles (**Supplementary Figure 2-4**). Neither global nor tail-based goodness-of-fit tests supported departure from the multiplicative model for any risk factor, for

both overall and ER-positive breast cancer (**Supplementary Table 4**). Goodness-of-fit tests were not performed for ER-negative breast cancer due to relatively small sample size.

Lack of evidence for substantial departure from the multiplicative model between the PRS_{313} and risk factors in this large study population implies that the absolute risk associated with each classical risk factor is expected to be larger for women with high versus low polygenic risk[5]. This is illustrated by our projections, which show that the lifetime risk due to classical risk factors was higher with a wider spread among women that are at a higher than lower risk due to genetic factors (PRS_{313} and family history) (**Figure 1a**), and recently for BMI stratified by familial risk[18]. The average lifetime risk of women due to all classical risk factors in the lowest and highest deciles of the genetic risk are 21.9% and 4.4%, with a risk difference between these two deciles of 17.5%. The corresponding risk difference between these two deciles due to the subset of modifiable risk factors is 16.5% (**Figure 1b**). However, the absolute risk projections shown in Figure 1 should be viewed with caution since they assume perfect model calibration. In addition, these absolute risk models require independent validation.

Our analyses using the current PRS_{313} are based on a sample size three times larger than that used in previously published BCAC analyses[9], although the dataset for ER-negative breast cancer is still limited. Our previous work on the PRS_{313} development[8] and the current analyses are based on European ancestry and may not be generalizable to other populations, highlighting the need for additional studies in populations of non-European or mixed ancestry.

Overall, the combined associations of the newly developed PRS_{313} and classical risk factors on breast cancer risk are well explained by a multiplicative model, except for family history. This suggests that preventive strategies aimed at modifying individual risk factors could

have stronger impact on risk reduction in women at higher genetic risk. Further, this is an important finding that will also inform the development of overall and ER-specific risk prediction models in future.

References

1. Rebbeck TR, Burns-White K, Chan AT*, et al.* Precision Prevention and Early Detection of Cancer: Fundamental Principles. Cancer Discov 2018;8(7):803-811.
2. Roberts MC. Implementation Challenges for Risk-Stratified Screening in the Era of Precision Medicine. JAMA Oncol 2018;4(11):1484-1485.
3. Cintolo-Gonzalez JA, Braun D, Blackford AL*, et al.* Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications. Breast Cancer Res Treat 2017;164(2):263-284.
4. Choudhury PP, Wilcox AN, Brook MN*, et al.* Comparative validation of breast cancer risk prediction models and projections for future risk stratification. J Natl Cancer Inst 2019; 10.1093/jnci/djz113.
5. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. J Natl Cancer Inst 2014;106(11).
6. Lee A, Mavaddat N, Wilcox AN*, et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 2019; 10.1038/s41436-018-0406-9.
7. Pharoah PD, Antoniou AC, Easton DF*, et al.* Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med 2008;358(26):2796-803.
8. Mavaddat N, Michailidou K, Dennis J*, et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet 2019;104(1):21-34.
9. Rudolph A, Song M, Brook MN*, et al.* Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. Int J Epidemiol 2018;47(2):526-536.
10. Maas P, Barrdahl M, Joshi AD*, et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA Oncol 2016;2(10):1295-1302.
11. Mavaddat N, Pharoah PD, Michailidou K*, et al.* Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 2015;107(5).
12. Michailidou K, Hall P, Gonzalez-Neira A*, et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45(4):353-61, 361e1-2.
13. Amos CI, Dennis J, Wang Z*, et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev 2017;26(1):126-135.
14. Michailidou K, Lindstrom S, Dennis J*, et al.* Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551(7678):92-94.
15. Milne RL, Kuchenbaecker KB, Michailidou K*, et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat Genet 2017;49(12):1767-1778.
16. Meisner A, Kundu P, Chatterjee N. Case-Only Analysis of Gene-Environment Interactions Using Polygenic Risk Scores. Am J Epidemiol 2019;188(11):2013-2020.
17. Song M, Kraft P, Joshi AD*, et al.* Testing calibration of risk models at extremes of disease risk. Biostatistics 2015;16(1):143-54.
18. Hopper JL, Dite GS, MacInnis RJ*, et al.* Age-specific breast cancer risk by body mass index and familial risk: prospective family study cohort (ProF-SC). Breast Cancer Res 2018;20(1):132.

Table 1: Odds ratios and 95% confidence intervals for multiplicative interactions between the 313 SNP-polygenic risk score (continuous) and classical risk factors of breast cancer, overall and by estrogen receptor (ER) status

		Case-control logistic regression method									Case-only linear regression method[#]					
		Overall breast cancer risk			ER-positive breast cancer risk			ER-negative breast cancer risk			Overall breast cancer risk		ER-positive breast cancer risk		ER-negative breast cancer risk	
Risk Factors	Controls	Cases	OR_{int} (95% CI)	p_{int}	Cases	OR_{int} (95% CI)	p_{int}	Cases	OR_{int} (95% CI)	p_{int}	OR_{int} (95% CI)	p_{int}	OR_{int} (95% CI)	p_{int}	OR_{int} (95% CI)	p_{int}
Age at menarche (per 2 years)	64087	52170	1.01 (0.99-1.03)	0.26	36820	1.01 (0.99-1.03)	0.29	8323	1.01 (0.98-1.04)	0.55	1.01 (1.00-1.02)	0.22	1.01 (0.99-1.02)	0.37	1.02 (0.99-1.06)	0.21
Ever parous (yes/no)	72552	59298	0.97 (0.93-1.00)	0.07	41858	0.98 (0.94-1.02)	0.35	9273	0.98 (0.92-1.05)	0.57	0.97 (0.94-1.00)	0.08	0.99 (0.96-1.03)	0.72	1.00 (0.92-1.09)	0.97
Number of children (1,2,3,≥4)*	61654	48786	1.00 (0.99-1.02)	0.96	34666	1.00 (0.99-1.02)	0.73	7552	0.99 (0.96-1.02)	0.53	1.01 (0.99-1.02)	0.38	1.01 (1.00-1.03)	0.11	1.00 (0.97-1.04)	0.90
Age at FFTP (per 5 years)*	53042	41671	1.00 (0.99-1.02)	0.82	29601	1.00 (0.98-1.01)	0.68	6517	0.99 (0.96-1.02)	0.52	1.00 (0.98-1.01)	0.39	0.99 (0.97-1.00)	0.06	1.00 (0.97-1.03)	0.92
Breastfeeding (yes/no)*	37568	34199	1.02 (0.98-1.06)	0.44	24273	1.01 (0.96-1.05)	0.81	5548	1.01 (0.95-1.08)	0.74	1.02 (0.99-1.05)	0.17	1.02 (0.98-1.06)	0.36	1.02 (0.95-1.11)	0.55
Duration of breast feeding (per 12 months)*	26367	27741	1.00 (0.98-1.02)	0.71	19329	1.00 (0.97-1.02)	0.76	4669	0.99 (0.95-1.03)	0.57	1.01 (0.99-1.03)	0.32	1.01 (0.99-1.03)	0.57	0.99 (0.96-1.03)	0.77
Adult height (per 5 cm)	62414	54847	0.99 (0.98-1.00)	0.07	38730	0.99 (0.98-1.00)	0.04	8682	1.00 (0.98-1.02)	0.77	1.00 (0.99-1.01)	0.92	0.99 (0.98-1.01)	0.29	1.01 (0.99-1.03)	0.48
Premenopausal BMI (per 5 kg/m2)†	15610	12837	0.98 (0.95-1.00)	0.08	8354	0.99 (0.96-1.02)	0.48	2333	0.95 (0.91-1.00)	0.04	0.97 (0.94-1.00)	0.02	1.00 (0.96-1.03)	0.77	0.95 (0.89-1.01)	0.10
Postmenopausal BMI (per 5 kg/m2)§	46137	37088	1.01 (0.99-1.02)	0.49	27305	1.01 (0.99-1.02)	0.39	5260	1.01 (0.99-1.04)	0.36	1.01 (1.00-1.02)	0.29	1.01 (1.00-1.03)	0.08	0.99 (0.96-1.02)	0.45
Ever use of oral contraceptives (yes/no)	56768	44979	1.01 (0.98-1.04)	0.63	31640	1.02 (0.98-1.05)	0.36	7061	1.02 (0.97-1.08)	0.42	0.99 (0.97-1.02)	0.45	1.00 (0.97-1.02)	0.75	1.01 (0.95-1.08)	0.73

Current use of EPT (yes/no)[§,‖]	20896	19047	1.07 (1.01-1.14)	0.02	14465	1.06 (0.99-1.13)	0.08	2761	1.05 (0.92-1.19)	0.49	1.00 (0.96-1.04)	0.93	0.98 (0.93-1.03)	0.32	1.04 (0.91-1.18)	0.59
Current use of Estrogen-only therapy (yes/no)[§,‖]	20716	18716	0.97 (0.91-1.03)	0.33	14201	0.96 (0.90-1.03)	0.28	2733	1.06 (0.94-1.20)	0.37	0.96 (0.91-1.01)	0.09	0.94 (0.89-0.99)	0.03	1.08 (0.95-1.23)	0.26
Alcohol consumption (per 10g/day)	16851	14484	1.00 (0.97-1.02)	0.75	10253	0.98 (0.96-1.00)	0.07	2259	1.06 (1.01-1.11)	0.03	1.00 (0.99-1.02)	0.71	0.99 (0.97-1.01)	0.19	1.04 (1.00-1.08)	0.06
Current smoking (yes/no)[¶]	56308	43303	1.04 (1.00-1.08)	0.07	30486	1.05 (1.00-1.10)	0.03	6813	1.05 (0.97-1.13)	0.25	1.02 (0.98-1.05)	0.42	1.02 (0.98-1.06)	0.40	1.03 (0.95-1.11)	0.52
Pack-years of smoking (per 10 pack-years)[**]	15990	11766	0.99 (0.98-1.01)	0.43	8268	0.99 (0.97-1.01)	0.19	1778	0.99 (0.96-1.02)	0.67	1.00 (0.99-1.01)	0.97	1.00 (0.99-1.01)	0.99	1.00 (0.97-1.03)	0.97
Family history in a first-degree relative (yes/no)	50955	42024	0.93 (0.89-0.96)	0.00003	28909	0.93 (0.90-0.97)	0.0008	6921	0.93 (0.87-0.99)	0.03	--	--	--	--	--	--

OR_{int}: Interaction odds ratio (per SD of PRS_{313}), CI: confidence intervals, SNP: single nucleotide polymorphisms, FFTP: First full-term pregnancy, BMI: Body mass index, MHT: Menopausal hormonal therapy, EPT: Estrogen-progesterone therapy.
Number of cases are same for case-control and case-only method
[#] The case-only analyses do not provide additional evidence to case-control analyses
Models are adjusted for reference age, study and ten ancestry-informative principal components
-- PRS and family history are not independent therefore, case-only analyses were not conducted for family history.
[*] Among parous women
[†] Among premenopausal women
[§] Among postmenopausal women
[‖] Models used to assess association with the use of MHT have been further adjusted for former use of any MHT, and use of other MHT preparations than the MHT preparation of interest
[¶] Models used to assess association with current smoking have been further adjusted for former smoking
[**] Among ever smoked

Figure 1: Distribution of absolute lifetime risk explained by a) all classical risk factors, b) modifiable classical risk factors within decile categories of genetic risk, due to 313-variant polygenic risk score (PRS) and family history, for 50-year old White non-Hispanic women in the United States before 80 years.

The solid horizontal lines represent the mean risk within each decile, while the dashed horizontal line across the plot represents the population lifetime mean risk (10.9%). Lifetime risk is estimated using the iCARE-BPC3 model and refers to absolute risk from age 50 to 80 years. The genetic component includes the 313-variant polygenic risk score and breast cancer family history. The classical risk factor component includes following risk factors: age at menarche, age at menopause, parity, age at first birth, height, body mass index, alcohol intake, smoking status, ever and current use of hormone replacement therapy (HRT), and HRT type among ever users. The modifiable classical risk factor component includes BMI, ever or current use of HRT, smoking status, and alcohol consumption. Outliers defined as points beyond 1.5 times the interquartile range below the first quartile or above the third quartile were excluded from the plot.

Funding and Acknowledgements

Funding

This work was supported by following funding agencies.

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and

Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136);

Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The CTS was supported by the California Breast Cancer Act of 1993, the California Breast Cancer Research Fund (contract 97-10500) and the National Institutes of Health (R01 CA77398, K05 CA136967, UM1 CA164917, and U01 CA199277). Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. HAC receives support from the Lon V Smith Foundation (LVS39420). The coordination of EPIC is financially supported by the European

Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the

German Cancer Research Center (DKFZ). The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. The MCBCS was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.The MEC was

support by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. The NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The Carolina Breast Cancer Study was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. PROCAS is funded from NIHR grant PGfAR 0707-10031. Prof. Dr. D Gareth Evans is supported by the NIHR Biomedical Research Centre in Manchester

(IS-BRC-1215-20007). The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. SEARCH is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER). The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01.

Acknowledgements

The authors thank all the researchers, clinicians, technicians, and administrative staff involved in the Breast Cancer Association Consortium. The authors would also like to thank all study participants, researchers, clinicians, technicians and administrative staff who participated in the parent studies (ABCFS, ABCS, ABCTB, AHS, BBCC, BCEES, BCINIS, BREOGAN, CBCS, CECILE, CGPS, CNIO-BCS, CPSII, CTS, EPIC, ESTHER, GENICA, GESBC, KARMA, kConFab /AOCS, LMBC, MARIE, MCBCS, MCCS, MEC, MISS, MMHS, NBHS, NC-BCFR, NCBCS, NHS, NHS2, OFBCR, PBCS, PKARMA, PLCO, PROCAS, SASBAC, SBCS, SEARCH, SISTER, SMC, UCIBCS, UKBGS, USRT, WHI) and have enabled this work to be carried out. The authors thank the staff of the Center for Genetic Epidemiology Laboratory, the CNIO genotyping unit, the McGill University and Génome Québec Innovation Centre, and the Mayo Clinic Genotyping Core Facility. CPSII acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The GENICA Network thanks Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany, German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - EXC 2180 – 390900677, Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany, Institute of Pathology, University of Bonn, Germany, Molecular Genetics of Breast Cancer, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany, and Institute of Occupational

Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany. KARMA and SASBAC studies thank the Swedish Medical Research Counsel. kConFab/AOCS wishes to thank all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. NHS and NHS2 would like to thank the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. PROCAS thanks NIHR for funding. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre.

Conflict of Interest: none declared